International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 3 Issue 3, March 2014

TGI Modules for Social Tagging System

Mr. Tambe Pravin M.
M.E. (2nd Year) Department Of Computer Engineering
SPCOE, Otur
Pune, India

Prof. Shamkuwar Devendra O.
Department Of Computer Engineering
SPCOE, Otur
Pune, India

***Abstract*— A Tag grouping and intimidation model (TGI) is presented for item recommendation in social tagging system. This model consists of three distinctive steps, in each of which important innovative elements are invented. In which, the content information is exploited to propagate tags between similar item, in order to solve sparsity and cold start problems. The sparsity is further handled, by generating tag clusters and revealing topics. By experimentally prove that a expert tags can improve the performance of quality and the associations among triplet's users, topics, and items are revealed by the TF technique of high order singular value decomposition. The TGI model finds problems of real-world applications, which produce noise and decrease the quality of recommendations.**

***Index Terms*— Expert tagging, social tagging, recommender systems, relevance feedback, Content based information retrieval.**

I. Introduction

SOCIAL tagging [11] is a process that allows users to set different items like photos, songs or web sites in order to their sharing, discovery and retrieval. These annotations are in the form of social tags or free keywords, through which users can express their personal opinion about different items. This is very important; since for the complex multifaced information of items like images, videos, etc. can be exploited and social tagging systems are generate personalized recommendations. Meanwhile, the users having the same tagging behavior to get similar recommendations. Therefore, such recommender system plays a very vital role as "item collaborative filtering in social tagging systems". The several important problems produce noise and reduce the recommendation accuracy. These problems are as following:

• Due the free natural tags, polysemy, and synonymy are very common problems, since tags are subject to multiple interpretations.
• "Learning tag relevance" is the problem of user tags are comparing to visual content.
• The "cold start" problem [14] refers to that users are participating rarely in the tagging process, so there are very few tags for the recommendation.
• Sparsity in social tagging systems affects the recommendation accuracy and decrease the quality. Specifically, since item recommendation in social tagging is collaborative, high accuracy is achieved when users having the same tagging behavior. However, this is very difficult in the real-world applications. A huge amount of social tagging data is generated with users and therefore the quality of recommendations is decreased.
• With the introduction of tags, the usual binary relation between users and items turns into a ternary relation between users, items and tags. This ternary relation is mapped to a tripartite network [12] which should be considered by recommender systems. i.e. users, tags, and items.

A. Background and Related Work

The expert tagging is a possible solution for the problems, which usually on a small number of domain experts, who annotate resources on structured vocabularies. The main advantages of using experts' opinions are: 1) the resulting well tag vocabulary and 2) the accurate annotations. However, the disadvantages are 1) the time required for the manual annotation is very large, and 2) the limited vocabulary.

In order to capture the ternary relation among users, tags, and items in social tagging systems, based on tensor factorization (TF) techniques [6]. Such methods are able to 1) solve problems like polysemy and synonymy; 2) preserve and maintain the ternary relation; 3) the latent associations among users, tags, and items; 4) reduce the noise in social tagging systems; 5) provide more accurate recommendation and 6) generate efficient recommendations in terms of run times, since the recommendations only depend on the smaller factorization dimensions after applying the TF method. The works do not solve sparsity, and the cold start problem, thus, the information is exploited by performing tag propagation in similar items. However, the main disadvantage of the method is that the tag propagation cannot be uncontrolled. More precisely, by allowing extensive tag propagation, the noise that may affect the quality of recommendations due to the irrelevant tag to items. However, the problem in this approach is the "learning tag relevance".

Since tag propagation between similar items can be performed only if they having the same concept. Continuously, a lot of work has been conducted by relevance feedback recommender systems [15]. Such approaches identify items that belong to the same concept while the tag dimension is always omitted. The item recommendation in social tagging systems is filtering based; the quality of recommendations is directly affected by sparsity. This is happen due to the fact that high recommendation accuracies are achieved. However, in the real-world applications, users have similar tagging behavior upon topics. The different methods aim at tag clustering, in which tags are clustered [9], [10]. However, these methods do not handle the problems of sparsity and "cold start," and generate inaccurate result.

B. Contribution and Layout

The Tensor Factorization and tag Clustering (TGI) model is for item recommendation in social tagging systems. The TGI model consists of three steps, and through each and every step, the problems are successfully solved. The first step of the system involves tag propagation by exploiting content, so that sparsity and "cold start" problems can be handled. The relevance feedback mechanism used in order to perform tag propagation between similar items if they belong to the same concept. Continuously, the TGI model is able to propagate less noisy tags. The TGI model is validated in an image collection and the method is not sufficient for image recommendation based on social tags, which increase the accuracy of recommendations.

The next step of the TGI model is tag clustering or grouping in order to reveal topics and identify the users in these topics. By performing so, the sparsity problem is solved, by transforming tags to tag clusters, and increase the accuracy of recommendations. The design method evaluated by performing two different tag clustering algorithms. The first one presented it takes into account the triplet network, and the second one is an accepts the *K*-Means in the social tagging systems for tag clustering. Through which, we conclude to the optimal number of tag clusters for both tag clustering methods, without considering the information of image classification. The $tf \cdot idf$ weighting scheme is followed to calculate users' interests and identify image relations to topics.

The last step of the analyzed method is based on the TF technique called high order singular value decomposition (HOSVD). The complexity of HOSVD is minimized by reducing the number of tags to tag clusters. The innovation in this step by exploiting HOSVD the associations among users, topics and images are revealed and exploits content to solve sparsity.

II. LITERATURE SURVEY

The related works are splited into four parts: 1) recommended systems based on the relevance feedback; 2) collaborative filtering in social tagging network; 3) tag clustering methods for topics and generating recommendations; and 4) exploiting expert tagging.

A. Recommender Systems Based on Relevance Feedback

Relevance feedback approaches have been used in several recommender systems try to recommend items similar to given user used in the past. The basic process of recommendation includes matching of the attributes and interests are stored, in order to recommend to user new interesting items.

The main difference between relevance feedback-based recommender systems and collaborative-based recommender systems in social tagging network is that relevance feedback-based recommendation systems try to recommend items similar to those user used in the past while systems about collaborative recommendation to identify users whose preferences are similar to those recommend items.

B. Collaborative Filtering in Social Tagging Systems

However, the user dimension in the triplet social network was eliminated along with the respective personalized recommendation. A generic method that allows tags to be incorporated to traditional, by reducing the three-way correlations to two-way. This technique of latent semantic analysis (LSA) [8] has been used for solution the problems of synonymy, polysemy, and noise in social tags. This technique became a popular for implementing filtering in social tagging systems. Nevertheless, the matrix factorization techniques ignore the three-way correlations between users, items, and tags.

The requirements of collaborative tagging systems such as the ternary relation, new approaches have been design. The ternary association among users, items, and tags are the effective tag recommendation using the tensor decomposition technique.

C. Tag Clustering for Generating Recommendations

The tag clustering methods have been design in the literature survey of recommender systems. Clustering is a grouping and offline process that is performed independently. The tag clusters form the basis of the recommendation and tag graph was built based on the tags in items. In [9], [10], several clustering techniques were evaluated *K*-Means and hierarchical clustering. A recommendation system based on hierarchical clustering of the tag, using user profiles and tag clusters results.

D. Expert Tagging

In [7], to find out expert tagging system in enterprises. In other systems [4] users were automatically classified into the expert tagging according to the preferences. For all the works, the ternary relation in the social tagging system was omitted and the quality was reduced. The expert tagging system usually operated on a small number of domain experts. Expert tagging provide the objective that cover multiple aspects. The main advantage of using experts is the resulting well vocabulary and accurate annotations. However, the main disadvantages are it requires very large time and limited vocabulary.

III. SYSTEM ARCHITECTURE

A. Overview and Architecture

In this section, the provided detailed description of the design TGI model. The social tagging data is a collection of triplets of users, tags, and images (*Ui, Tj, Ik, w)*, where the weight *w* that user *Ui* will tag item *Ik* with tag *Tj*.

The architecture of our system is composed of several components. They are as follows:

- RMTF is the ranking model tensor factor for the purpose of finding ranking of item. After getting response from user to tagging image it process for calculate ranking on the basis of its review and other status.
- User Specific Topic occurs after RMTF on the basis of interest of the user in particular item.

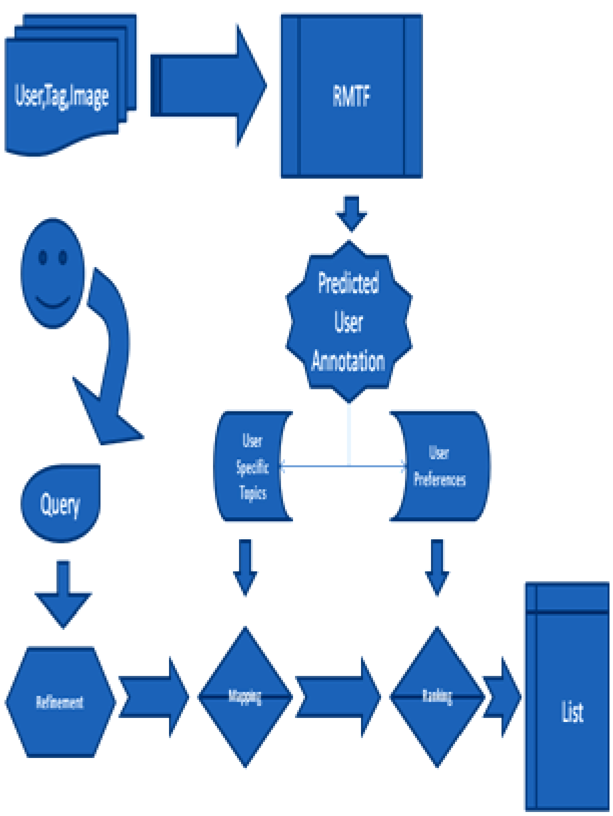

Fig. 1. System Architecture

- User Preference module checks the image relationship with the topic.
- Refinement used to checks whether user enters query is under different refinement conditions.
- Mapping/Ranking is the finite output given by user specific topic and user preferences respectively.
- List/Histogram displays the list is generated after ranking at front end and histogram at back end.

B. Tag Propagation and Architecture for Complex Words

The first step of the TGI model operates tag propagation based on a relevance feedback mechanism to propagate between similar images. Firstly, a descriptor vector DV is extracted from each and every image called "color and edge directivity descriptor" (CEDD). In particular, CEDD plays an important role in the color and texture information of images in a histogram, as follows: first, the image is separated in a number of blocks in order to extract the color information. A fuzzy rules used for the extraction of a histogram [5]. Furthermore, the 20 rules are applied to a three-input fuzzy system to generate a 10-bin quantized histogram.

The complex words plays a very important role in the system as per as different tags are getting concern. After entering any input query by the user dictionary checks all its propagated information like stop words, abbreviation, synonyms and tokenizer for preference processing.

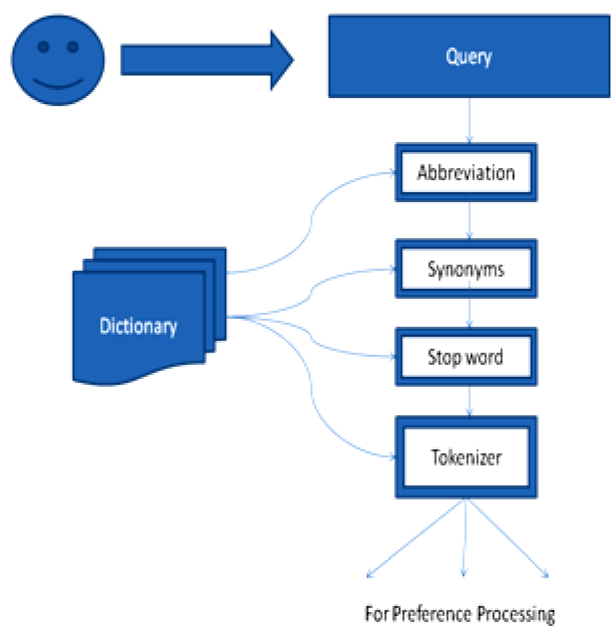

Fig. 2. Complex Words

Likewise, the following is the bucket model for searching an exact element from triplet tag, user, and image to getting complex words.

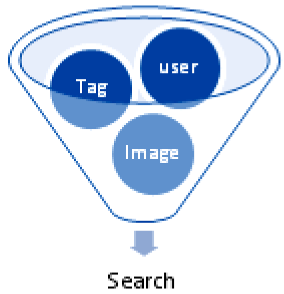

Fig. 3. Input/output Complex Words

IV. ALGORITHM

Algorithm 1. The algorithm for Complex Words

```
q =get query
D = initialize dictionary
for each(word in q )
{
if(word ∩ D)
{
word = d1
}
else
{
continue
}
for each(word in synonyms s )
{
if(word ∩ s)
{
word = s1
}
else
{
continue
}
for each(word in stop st)
{
if(word ∩ st)
{
Word = st1
}
else
{
continue
}
// tokenize string
q[] = tokenize(q,' ');
for each qtemp € q[i]
forward for preference search
end of for
```

The complex words plays a very important role in the system as per as different tags are getting concern. After entering any input query by the user dictionary checks all its propagated information like stop words, abbreviation, synonyms and tokenizer for preference processing. It is done through above mentioned algorithm by triplets like tag, user, and images.

A. Advantages

- A ranking based tensor factorization model named RMTF is analyzed to predict users' annotations to the images.
- To better represent the query-tag relationship, the builded user-specific topics and map the queries as well as the users' preferences onto the learned topic spaces.
- The resulting well agreed tag vocabulary.
- Increase quality and accuracy.
- TGI model includes tag propagation and relevance feedback mechanism.
- Reduce Noise.

B. Applications

- Item/Image classification and retrieval.
- Compression.
- Clustering.
- Material identification.
- Spectral Imaging (SOI) and Tensor analysis with NTF.

V. CONCLUSION AND FUTURE SCOPE

The Tag Grouping and Intimidation (TGI) model for item recommendation in social tagging network is a tag grouping and intimidation model, since from which each of its steps that handle several issues that exist in social tagging networks, which produce noise and decrease the accuracy and quality of system. The content information is exploited to propagate tags between similar items to solve the sparsity and cold start problems and the sparsity is further handled by generating tag clusters, triplets are revealed by TF technique.

ACKNOWLEDGMENT

We would like to thank Prof. Chandre P.R. for fruitful discussions and helpful comments on this work.

REFERENCES

[1] J. Ahn, P. Brusilovsky, J. Grady, D. He, and S. Y. Syn, "Open user profiles for adaptive news systems: Help or harm?" in *Proc. Int. Conf. WWW*, 2007, pp. 11–20.

[2] M. Balabanovic and Y. Shoham, "Fab: Content-based, collaborative recommendation," *Commun. ACM*, vol. 40, no. 3, pp. 66–72, 1997.

[3] G. Begelman, P. Keller, and F. Smadja, "Automated tag clustering: Improving search and exploration in the tag space," in *Proc. Collab.* Web Tagging Workshop, WWW, 2006.

[4] T. Bogers, W. Thoonen, and A. Van Den Bosch, "Expertise classification: Collaborative classification vs. automatic extraction," in *Proc. 17th SIG* Classif. Res. Workshop, Austin, TX, 2006.

[5] S. Chatzichristofis and Y. Boutalis, "A hybrid scheme for fast and accurate image retrieval based on color descriptors," in *Proc. IASTED Int. Conf.* Artif. Intell. Soft Comput., 2007, pp. 280–285.

[6] P. Drineas and M. W. Mahoney, "A randomized algorithm for a tensor based generalization," *Linear Algebra Appl.*, vol. 420, no. 2/3, pp. 553–571, 2007.

[7] S. Farrell and T. Lau, "Fringe contacts: People-tagging for the enterprise," in Proc. Collab. Web Tagging Workshop WWW, 2006.

[8] G. Furnas, S. Deerwester, and S. Dumais, "Information retrieval using a singular value decomposition model of latent semantic structure," in *Proc. ACM SIGIR Conf.*, 1988, pp. 465–480.

[9] J. Gemmell, A. Shepitsen, B. Mobasher, and R. Burke, "Personalization

in Folksonomies based on tag clustering," in *Proc. Intell. Tech. web Pers.* Recommender Syst. Workshop, 2008, pp. 37–48.
[10] J. Gemmell, A. Shepitsen, B. Mobasher, and R. Burke, "Personalizing navigation in folksonomies using hierarchical tag clustering," in *Proc. Int.* Conf. Data Warehousing Knowl. Discov., 2008, pp. 196–205.
[11] S. Golder and B. Huberman, "The structure of collaborative tagging systems," in *Proc. CoRR*, 2005, pp. 1–8.
[12] H. Halpin, V. Robu, and H. Shepherd, "The complex dynamics of collaborative tagging," in *Proc. Int. Conf. WWW*, 2007, pp. 211–220.
[13] Z. He, A. Cichocki, S. Xie, and K. Choi, "Detecting the number of clusters in n-way probabilistic clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2006–2021, Nov. 2010.
[14] J. Herlocker, J. Konstan, L. Terveen, and J. Riedl, "Evaluating collaborative faltering recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53, Jan. 2004.
[15] S. C. Hoi, M. R. Lyu, and R. Jin, "A unified log-based relevance feedback scheme for image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 4, pp. 509–524, Apr. 2006.

